# 联邦数据库系统

张　兵　张荣肖　潘玉平　　(中科院计算中心)

**摘要**：本文描述了在多数据库环境下异体结构的独立数据库管理系统中联邦概念的应用、系统开发过程和使用的方法。并介绍了一些实例系统。

在当今信息社会的时代，信息处理正在由集中式过渡到分布式，由于计算机和通讯技术的发展，人们逐步要求从相互独立运行多个数据库中获取信息，对这类问题的研究，在技术上称为异质数据库领域(Heterogeneous Database area)，建立联邦数据库系统(Federated database system)是解决这类应用问题的一种方法。通常称相互独立运行的数据库系统为单元数据库系统(Component DBS)。它是原来已存在的、在局部地区已操作运转的数据库系统，并且作为联邦数据库系统的一个成份。

联邦数据库系统的概念是由 Hammon 和 Mclcod(1979)及 Heimlighon 和 Mcleod(1985)提出的，它涉及到很广泛的需求范围，如联网信息服务，商业或金融组织，政府办公自动化，科学工程计算和环境等。

本文描述了联邦数据库系统的概念及应用，第一节讲述了 FDBS 的特征与分类，第二节说明了当前 FDBS 的参考结构，第三节是关于 FDBS 的开发方法，第四节是关于 FDBS 的操作，第五节介绍了国外应用实例。

## 一、联邦数据库系统的特征与分类

所谓联邦数据库系统(FDBS)是一个彼此协作却又相互独立的单元数据库(CDBS)的集合，它将单元数据库系统按不同程度进行集成，对该系统整体提供控制和协同操作的软件叫做联邦数据库管理系统(FDBMS)，一个单元数据库可以加入若干个联邦系统，每个单元数据库系统的 DBMS 可以是集中式的，也可以是分布式的，或者是另外一个 FDBMS.

### 1.联邦数据库的特征

联邦数据库系统(FDBS)最重要的特征就是：一个单元数据库系统在继续本地操作的同时可以参加联邦系统的活动。单元 DBS 的集成可以由联邦系统的用户来管理，也可由联邦系统的管理员和单元 DBS 的管理员来共同管理，整体系统集成的程度取决于联邦系统用户的要求、取决于加入联邦系统并共享联邦系统数据库的单元 DBS 的管理员的要求。

包含多种数据库系统的 FDBS，其特征还体现在分布，异质和独立这三个方面：

(1)分布：数据可按不同方式在多个数据库之间分布，这些数据库可通过通信系统关联。

(2)异质：许多类型的异质来源于技术上的不同，如硬件、系统软件和通讯系统的不同，可按 DBMS 和源于数据语义的不同来划分。$DBMS_s$ 的异质是由于不同的数据模型提供不同的结构基元或者支持不同的约束条件，以及查询语言不同；当涵义、解释、对同种或相关数据的使用出现不一致时，出现语义上的异质。

(3)独立：管理不同 $DBS_s$ 组织实体通常都是独立的，主要表现在；

①设计独立：单元 DBS 在数据管理、数据元素表达、数据的概念化和语义解释，数据管理的约束条件、系统功能、同其它系统的关联和共享、以及工具等方面设计自主。

②通讯独立：单元 DBMS 决定是否同其它 DBMS 通讯。及何时和怎样回答其它 DBMS 的请求。

③执行独立：单元 DBMS 执行本地操作而不受外部干扰，以及决定外部操作执行顺序的能力。

④相关独立：单元 DBS 决定是否同其它系统共享其功能和资源以及在多大程度上共享.包括同联邦关联或解除关联的能力及单元 DBS 加入一个或更多联邦的

能力.

**2.Multi-DBMS 和 FDBS 的分类**

多数据库系统(MDBS)支持在多单元 DBMS 上的操作。单元 DBS 由单元 DBMS 管理,在 MDBS 中的单元 DBS 可能是集中式或分布式的。

如果单元 $DBS_S$ 的 $DBMS_s$ 都一样,那么该 $MDB_s$ 即为同类的,否则称为异质(heterogeneous)的 MDBS。

基于单元 $DBS_s$ 的独立性,$MDBS_s$ 划分为非联邦数据库系统和联邦数据库系统.前者是由并非独立的单元 $DBS_s$ 集成的,所有操作统一完成,仅是一层管理,不区分本地与非本地用户.在用户看来,逻辑上象一个分布式 DBS.而联邦数据库系统是由独立并加入联邦的,允许部分有控制地共享其数据的单元 $DBS_s$ 组成的.在联邦结构中不存在集中控制,由单元 $DBS_s$ 控制对其数据的访问.在保护单元 $DBS_s$ 的独立性,继续其现存应用程序执行的同时,FDBS 支持本地和全局操作,可以提供有控共享。

由谁管理联邦和单元怎样集成,基于这些问题,$FDBS_s$ 可划分为松散耦合和紧耦合前者要求用户创建和维护联邦,联邦系统不强加控制.而后者则要联邦和其管理员来创建和维护联邦,并主动控制对单元 $DBS_s$ 的访问。

## 二、$FDBS_s$ 的参考结构

FDBS 中的基本成份——数据、数据库、命令、处理程序,模式和映射等,按不同方式组合可产生不同的数据管理结构。图 1 所示是 FDBS 的参考结构,它将处理程序划分为四类——每类都针对数据操纵命令和访问的数据完成不同功能;并为了处理 FDBS 的分布性,异质性和独立性,提出了不同于集中式的标准三级模式结构(概念模式、内模式、外模式)的五级模式结构。

**1.处理器**

(1)变换处理器:将命令从一种语言形式翻译成另一种语言形式,或将数据从一种格式变换成另一种格式。支持数据模型透明,从而掩盖了查询语言和数据结构上的差异。

(2)筛选处理器:限制能够传送到另一个处理器的命令和相关数据。与其关联的是映射,描述命令和数据上的约束。

(3)构造处理器:划分与/或复制某处理器发来的操作,使之能被两个或更多处理器接收的操作;也可将若干处理器产生的数据合并成一个单数据集供某个处理器使用。支持定位、分布和复写透明,主要完成模式集成、协商、查询和全局事务管理等任务。

(4)存取控制器:接收命令并在数据库上执行来产生数据。

**2.模 式**

(1)局部模式:是单元 DBS 的概念模式,用单元 DBMS 的本机数据模型表达。

(2)单元模式:将局部模式转译为 FDBS 的所谓规范或公共数据模型(Canonical or Common Data Model-CDM)的数据模型。

(3)输出模式:一个输出模式代表一个单元模式的子集,包括存取控制信息。FDBS 可用之。

(4)联邦模式:是多个输出模式的集成。

(5)外部模式:是为某个用户与/或应用程序或者某类用户/应用程序定义的模式。

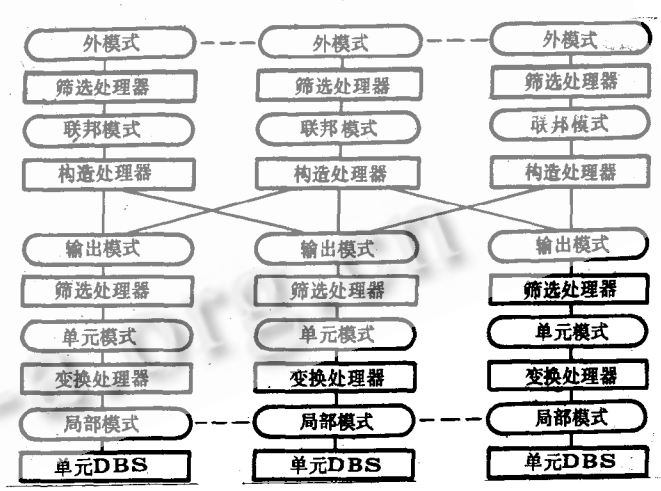


图 1 FDBS 的系统结构

## 三、联邦数据库系统的开发

用联邦方案来管理分布式数据,不必改变或废除现存的应用程序或完全改变信息管理的组织机构,支持现存数据库的受控集成,方便新的应用程序和数据库的加入.

**1.联邦数据库系统的开发过程**

FDBS 是通过一组相关联的 DBSS 的逐步集成发

展的，随着新单元数据库的增加和现存数据库的修改，FDBS 也随之发展。这种发展分三阶段：

(1)预集成：当数据驻留在文件中不被任何 DBMS 管理，但联邦用户又要访问时，可将文件移到某个 DBMS 中或者扩充文件系统使之支持类 DBMS 特征。

(2)开发联邦数据库系统：涉及到单元，输出、联邦和外部模式的建立，不同模式之间的映射定义以及相关处理器的形成。

(3)联邦数据库系统操作：涉及到用 FDBMS 管理和操纵多个集成的数据库。

这些阶段不必遵守从一个阶段到下一个阶段的顺序，每个阶段可以执行数次，可以重新访问前一阶段，也可修改前一阶段的结果。

**2.联邦数据库系统的开发方法**

(1)自底向上的开发方法：主要是集成现存的单元数据库来开发一个新的 FDBS，该方法适于向 FDBS 添加一个新的单元数据库。图 2 所示过程如下：

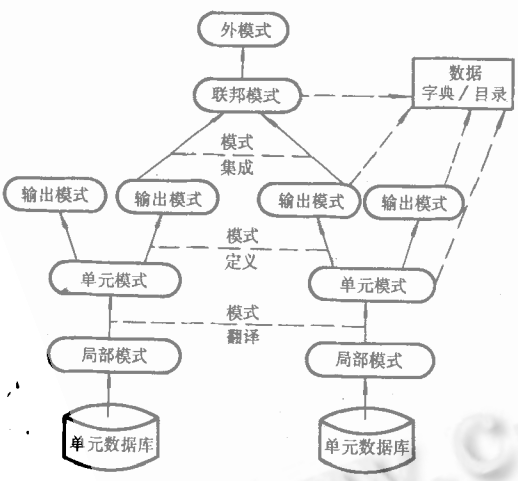


图 2　自底向上的 *FDBS* 开发过程

①翻译模式：将单元数据库的局部模式翻译成用 CDM 表达的单元模式；在两种模式的对象之间产生映射；开发(或辨识)变换处理器，使之能将用单元模式表述的命令变换成用相应的局部模式表达的命令。

②定义输出模式：根据单元模式定义输出模式(由相应的单元 $DBS_s$ 的管理员完成)。根据与联邦 DBA 的协议授权 FDBS 对单元 $DBS_s$ 的共享部分；开发(辨认)合适的筛选处理器。

③集成模式：选取某个欲被集成的输出模式集，将其集成并产生一个联邦模式；开发(或辨认)一个构造处理器，使之可以将用联邦模式表述的命令转换成用相应的输出模式表达的命令，包括生成含有适当分布信息的映射。

④定义外部模式：若有必要，可为每个／类联邦用户定义外部模式；构造(辨认)必需的筛选处理器或变换处理器。如果外部模式的数据类型同 CDM 的不一样，变换处理器还要进行模式翻译。

(2)自顶向下的开发方法：利用现存的 FDBS 开发新的应用程序时，有必要确定联邦模式是否能支持应用程序的数据需求。若不能就应该扩充某个联邦模式或建立一个新的联邦模式，或者扩充某个现存的单元数据库或建立一个新的单元数据库。此过程即为自顶向下，它是对传统分布式数据库设计的过程扩充。图 3 所示过程如下：

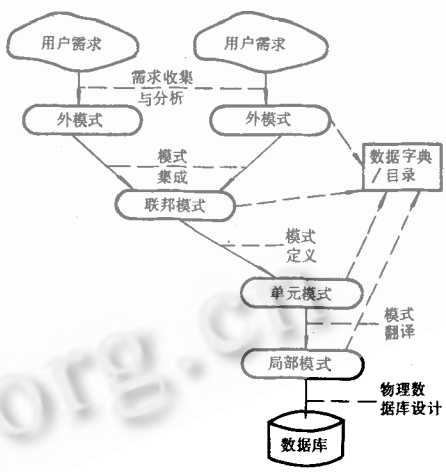


图 3　自顶向下的 *FDBS* 开发过程

①定义或修改外部模式：收集联邦用户需求并加以分析，定义新的或扩充现存的外部模式。

②分析模式：将外部模式同相关的联邦模式比较，以识别出早已存于联邦模式并因此而被 FDBS 支持的外部模式部分，若有部分不被 FDBS 支持，就需扩充或开发一联邦模式以包含这些部分——作为临时模式，需要一个或更多的单元数据库支持它。

③集成模式：将临时模式和联邦模式集成在一起，然后删除临时模式。

**3.联邦数据库系统的开发任务**

开发一个集中式或分布式 DBS 的工作跟开发 FDBS 的很有关系，只要少量修改就可适用。这里介绍与模式开发相关的很重要的四项任务：

(1)模式翻译：将用某种数据模型表达的模式映射到等价的用另一种数据模型表达的模式。

(2)存取控制：在图 1 的系统结构中，联系输出和单元模式的筛选处理器控制对单元 $DBS_s$ 的访问；联系外部和联邦模式的筛选处理器控制对联邦模式的访问。在单元和联邦 $DBA_s$ 之间应达成一致协议，如何控制单元 $DBA_s$ 对某些联邦用户保密而对另一些公开的数据，以及识别用户安全保密等。

(3)协议：联邦 DBA 管理联邦模式，单元 $DBA_s$ 管理定义在其管理的单元 $DBS_s$ 上的输出模式。联邦 DBA 和单元 $DBA_s$ 就输出模式的内容和输出模式上允许的操作应达成一致意见。这样就可在输出模式上定义联邦模式，以支持联邦用户。管理员之间的对话即为协议。

(4)模式集成：指的是将多个模式综合成一个单独的模式，一般可将其划分为预集成，比较、组织、合并和调整五步。

## 四、联邦数据库系统的操作

许多与分布式 DBMS 的操作有关的工作也与多数据库系统和 $FDBS_s$ 的操作有关，下面的四项任务是特别针对 FDBS 的：

**1.查询表达**

在集中式和分布式 DBMSS 中使用的查询语言适用于紧耦合的 FDBS。大多数的松散耦合数据库提供多数据库语言，允许联邦用户从多个单元 $DBS_s$ 存取数据，其提供的功能在集中式或分布式的 $DBMS_s$ 的数据操纵语言中是没有的。

**2.命令变换**

命令变换处理器将某种语言中的命令翻译成另一种语言中的命令。

**3.查询处理和优化**

松散耦合的 FDBS、FDBMS 几乎不支持查询优化，而在紧耦合的系统中可以实现广泛的查询优化。查询过程将对联邦模式的查询转换为在几种输出模式上的查询，并执行之。FDBMS 中的查询处理同分布式 DBMS 中的很相似。

**4.全局事务管理**

全局事务管理程序在允许对多数据库同时修改的时候，负责维护数据库的一致性。

## 五、实例介绍

**1.ADDS (Amoco Production Company,Research)**

Amoco 分布式数据库系统，即 ADDS 项目自 1993 年开始，主要是处理在集成分布于企业中的数据库时所出现的问题，应用程序需用到来自多个源地的数据，最初目标是简化 Amoco 内的分布式数据的存取和管理。

ADDS 基于关系式数据模型并使用一种扩展关系代数查询语言，同时也支持 ANSI SQL 语言标准的子集，按[sheth 和 Larso 1990] 所述，ADDS 是一个支持多联邦模式的紧耦合的联邦系统。该系统包括地理上分布的运行 VM 和 MVS 操作系统的主机以及运行 UNIX 操作系统的 SUN 和 Apollo 工作站，采用了统一的网络接口——网络接口设施(NIFTY)结构是 OSI 参考模型的扩展，可为使用不同物理通信网络的计算机系统提供统一可靠的接口。ADDS 保持局部数据库系统的独立性，不需修改局部 DBMS 软件。

**2.DATAPLEX(General motors Corporation)**

由于数据管理需求的不同，在加工工业中只好使用许多不同的数据库管理系统和文件系统。由于缺少有效的途径来共享这些异质的数据库，从而导致设计加工以及商务的效率不高，为此通用摩托公司开发了异质分布式数据库系统 DATAPLEX 原型，它与运行在 MVS 操作系统下的 IMS 的层次型 DBMS 相接，并与运行在 VMS 操作系统的 VAX 机上的 INGRES 的关系式 DBMS 相接。

DATAPLEX 也是紧耦合的多联邦模式的 FDBS。

**3.IMDAS(National Institute of Standards and Technology)**

在现代化的生产系统，工业自动化和计算机集成生产(CIM)的发展是很重要的。在大多数工业设备中，控制、设计和管理系统都运行于不同厂家的计算机系统和数据库系统上，它们独立设计，数据库重叠，对于同样的真实世界中的物体只是逻辑和物理表达不同。这些系统投资大，要想重新设计以取代之是不现实的。

集成化生产数据管理系统(IMDAS)的开发主要是

支持原型化的 CIM 环境——NBS 自动化生产研究设施(Automated Manufacturing Research Facility (AMRF)),一种测试小批量生产自动化和在制品测量的平台。目的主要是从许多系统存取数据,在使新的和更新过的应用程序存取数据的同时,与现存数据库中的现存应用程序协作,并协调在定位、表达和存取机制上的不一致。

IMDAS 是一紧耦合的联邦,只有单一的全局模式,集成数据模型是语义相关模型。支持分布式更新(事务管理)和分布式检索(查询管理)。

**4.其 它**

还有其它一些紧耦合的,支持多联邦模式的联邦系统。象 Ingres 公司的 INGRES 系统,允许用户访问分布数据库;Data Integration 公司的 Mermaid 系统,最早是在 1982 年开发的,主要是解决异质环境(硬件,操作系统,DBMS 等)下数据存取和集成的操作问题。

MULTIBASE(Xeror Advanced Information Techno logy)是 1980 年开发的紧耦合联邦系统,提供对多局部与联邦模式的定义,它是一个通用系统,并不局限于特殊领域,不改变现存数据库,也不干扰现存的应用系统,包含广泛的数据源,并提供统一的集成化接口,从已存的异质分布的数据库中检索数据。

Sybase 公司的 SYBASE 系统是松散耦合的联邦系统(协配操作系统),它希望尽可能开放结构,这样任何数据库,应用程序或服务都可集成到异质环境中的委托/服务器结构中,支持分布式操作,基于关系模式,基本的查询语言是 SQL。

## 六、结束语

建立联邦数据库系统是当前解决异质数据库领域中具体实用问题值得探讨的一种方法。当前的研究指出:通过建立标准,排斥各单元子数据库系统间的异质性,就使问题转化为探讨使标准一致的方法及其困难程度。在文献 3 中,讨论了当前标准化工作的进展情况。

**参考文献:**

*[1]"Federated Datadase Systems for Managing Distributed,Heterogeneous and Autonomous Databases" AMIT P.SHETH,JAMES A.LARSON*

*[2]"Heterogeneous distributed database Systems for Production use"GOMER THOMAS ,GLENN R.THOMPSON 等*

*[3]"Interoperability of Multiple Autonomous Databases"WITOLD LITWIN,LEO MARK, NICK ROUSSOPOULOS*